AMENDMENTS TO THE CLAIMS

1 (previously presented): A method of detecting a facial region within a video comprising the steps of:

(a) receiving a first frame of said video comprising a plurality of pixels;

(b) receiving a subsequent frame of said video comprising a plurality of pixels;

(c) calculating a difference image representative of the difference between a plurality of said pixels of said first frame and a plurality of said pixels of said subsequent frame;

(d) determining a plurality of candidate facial regions within said difference image based on a transform of said difference image in a spatial domain to a parameter space; and

(e) fitting said plurality of candidate facial regions to said difference image, where said difference image used for said fitting is free from being transformed as a result of step (d), to select one of said candidate facial regions.

2 (previously presented): The method of claim 1 further comprising the step of thresholding said difference image thereby removing values of said difference image less than a threshold value.

3 (previously presented): The method of claim 2 wherein said threshold value is a predetermined value and said removing values is setting said values of said difference image that are less than said threshold value to a selected value.

4 (previously presented): The method of claim 1 wherein said transform is a Hough transform.

5 (currently amended): ~~The method of claim 4 wherein said Hough transform is~~

~~$A(x_c, y_c, r) = A(x_c, y_c, r) + 1 \; \forall \; x_c, y_c, r \; \varepsilon \; (x-x_c)^2 + (y-y_c)^2 = r^2$.~~ A method of detecting a facial region within a video comprising the steps of:

(a) receiving a first frame of said video comprising a plurality of pixels;

(b) receiving a subsequent frame of said video comprising a plurality of pixels;

(c) calculating a difference image representative of the difference between a plurality of said pixels of said first frame and a plurality of said pixels of said subsequent frame;

(d) determining a plurality of candidate facial regions within said difference image based on a Hough transform of said difference image in a spatial domain to a parameter space of the form
$A(x_c, y_c, r) = A(x_c, y_c, r) + 1 \; \forall \; x_c, y_c, r \; \varepsilon \; (x-x_c)^2 + (y-y_o)^2 = r^2$; and

(e) fitting said plurality of candidate facial regions to said difference image, where said difference image used for said fitting is free from being transformed as a result of step (d), to select one of said candidate facial regions.

6 (previously presented): The method of claim 1 where said fitting of each of said candidate facial regions is based on a combination of at least three factors including, a fit factor representative of a fit of said candidate facial regions to said difference image, a location factor representative of the location of said candidate facial regions within said video, and a size factor representative of the size of said candidate facial regions.

7 (previously presented): The method of claim 1 further comprising the step of scaling said first frame and said subsequent frame of said video to reduce the number of said pixels of said first and subsequent frame prior to said calculating said difference frame.

8 (currently amended): ~~The method of claim 1 wherein said step of determining said plurality of said candidate facial regions and fitting said plurality of said candidate facial regions further comprises the steps of:~~ A method of detecting a facial region within a video comprising

the steps of:

(a) receiving a first frame of said video comprising a plurality of pixels;

(b) receiving a subsequent frame of said video comprising a plurality of pixels;

(c) calculating a difference image representative of the difference between a plurality of said pixels of said first frame and a plurality of said pixels of said subsequent frame;

(d) determining a plurality of candidate facial regions within said difference image based on a transform of said difference image in a spatial domain to a parameter space wherein said step of determining said plurality of said candidate facial regions and fitting said plurality of said candidate facial regions further comprises the steps of:

~~(a)~~(i) determining a set of candidate circles based on a Hough transform of said difference image;

~~(b)~~(ii) scoring said set of said candidate circles based on a combination of at least three factors including, a fit factor representative of the fit of said candidate circles to said difference image, a location factor representative of the location of said candidate circles within said video, and a size factor representative of the size of said candidate circles;

~~(c)~~(iii) selecting at least one of said candidate circles based on said scoring;

~~(d)~~(iv) generating at least one candidate facial region having an elliptical shape for each of said at least one of said candidate circles; and

~~(e)~~(v) scoring each of said candidate facial regions based on a combination of at least three factors including, a fit factor representative of the fit of a respective said candidate facial region to said difference image, a location factor representative of the location of said respective said candidate facial region within said video, and a size factor representative of the size of said respective said candidate facial region; and

(e) fitting said plurality of candidate facial regions to said difference image, where said difference image used for said fitting is free from being transformed as a result of step (d), to select one of said candidate facial regions.

9 (previously presented): The method of claim 8 wherein said generating at least one candidate facial region has a center of said elliptical shape located within a bounded region of potential locations having a greater vertical dimension than a horizontal dimension centered about the center of said respective said candidate circle.

10 (previously presented): A method of detecting a facial region within a video comprising the steps of:

(a) receiving a first frame of said video comprising a plurality of pixels;

(b) receiving a subsequent frame of said video comprising a plurality of pixels;

(c) calculating a difference frame representative of the difference between a plurality of said pixels of said first frame and a plurality of said pixels of said subsequent frame;

(d) determining a plurality of candidate facial regions within said difference frame; and

(e) fitting said candidate facial regions to said difference image to select one of said candidate facial regions based on a combination of at least two of the following three factors including, a fit factor representative of the fit of said candidate facial regions to said difference image, a location factor representative of the location of said candidate facial regions within said difference image, and a size factor representative of the size of said candidate facial regions

11 (previously presented): The method of claim 10 where said determining said candidate facial regions is based on a Hough transform of said difference image in a spacial domain to a

parameter space.

12 (currently amended): ~~The method of claim 11 wherein said Hough transform is~~ ~~$A(x_c, y_c, r) = A(x_c, y_c, r) + 1 \; \forall \; x_c, y_c, r \in (x-x_c)^2 + (y-y_c)^2 = r^2$.~~ A method of detecting a facial region within a video comprising the steps of:

(a) receiving a first frame of said video comprising a plurality of pixels;

(b) receiving a subsequent frame of said video comprising a plurality of pixels;

(c) calculating a difference frame representative of the difference between a plurality of said pixels of said first frame and a plurality of said pixels of said subsequent frame;

(d) determining a plurality of candidate facial regions within said difference frame based on a Hough transform of said difference image in a spatial domain to a parameter space wherein said Hough transform is $A(x_c, y_c, r) = A(x_c, y_c, r) + 1 \; \forall \; x_c, y_c, r \; \varepsilon \; (x-x_c)^2 + (y-y_c)^2 = r^2$; and

(e) fitting said candidate facial regions to said difference image to select one of said candidate facial regions based on a combination of at least two of the following three factors including, a fit factor representative of the fit of said candidate facial regions to said difference image, a location factor representative of the location of said candidate facial regions within said difference image, and a size factor representative of the size of said candidate facial regions.

13 (previously presented): The method of claim 10 further comprising the step of thresholding said difference image thereby removing values of said difference image less than a threshold value.

14 (previously presented): The method of claim 10 further comprising the step of scaling said first frame and said subsequent frame of said video to reduce the number of said pixels of said first and subsequent frame prior to said calculating said difference frame.

15 (previously presented): A method of determining sensitivity information for a video comprising the steps of:

(a) receiving a first frame of said video;

(b) receiving a subsequent frame of said video;

(c) determining a spatial location of a facial region within said video based on at least said first and said subsequent frame; and

(d) calculating a sensitivity value for each of a plurality of spatial locations within said video based upon both said spatial location of said facial region within said video in relation to said spatial plurality of locations and a non-linear model of the sensitivity of a human visual system's ability to perceive image detail at eccentric visual angles.

16 (previously presented): The method of claim 15 wherein the step of said calculating said sensitivity values is further based upon calculating an eccentricity versus image location in relation to a viewer of said video for said plurality of locations within said video.

17 (previously presented): The method of claim 16 wherein said calculating said sensitivity is further based upon a sensitivity versus eccentricity non-linear model of said human visual system.

18 (currently amended): The method of claim 16 wherein said eccentricity is derived according to the following,

$$\theta_E = \frac{180}{\pi} \tan^{-1}\left(\frac{\sqrt{\left(\frac{y - y_c}{y_R}\right)^2 + \left(\frac{x - x_c}{x_{UDR}}\right)^2} - 1}{V} \right)$$

where θ_E is said eccentricity, y is a vertical pixel position within said video, x is a horizontal

position within said video, x_c represents a horizontal component of a center position of an elliptical said facial region, y_c represents a vertical component of said center position of said elliptical said facial region, x_r represents a first elliptical radii of said elliptical said facial feature in a horizontal direction; y_r represents a second elliptical radii of said elliptical said facial feature in a vertical direction, and V represents a viewing distance.

19 (previously presented): The method of claim 15 wherein said sensitivity values are based upon the distance from the outer edge of said facial region to said plurality of locations within said video.

20 (previously presented): The method of claim 17 wherein said sensitivity versus eccentricity non-linear model is derived according to the following,

$$S = \frac{1}{1 + k_{ECC}\theta_E}$$

where S is representative of said sensitivity, k_{ECC} is a constant, and θ_E is representative of a non-linear contrast sensitivity function.

21 (previously presented): A method of encoding a video comprising the steps of:

(a) receiving a frame of said video consisting of a plurality of pixels;

(b) calculating sensitivity information for a plurality of locations within said frame of said video calculated based upon the sensitivity of a human visual system of a viewer perceiving image detail at eccentric visual angles of a particular region of said frame of said video, where said particular region of said frame is determined based upon the content of the frame itself; and

(c) encoding said frame in a manner that provides a substantially uniform apparent quality to perceiving detail at eccentric visual angles of said plurality of locations of said frame to said viewer when said viewer is observing said particular region of said frame of said video.

22 (previously presented): The method of claim 21 wherein said encoding of each of said plurality of locations of said frame of said video is based on a respective quantization value representative of a base quantization factor divided by said sensitivity information for a respective one of said plurality of locations in a manner that said encoding employs at least two different quantization values, where said plurality of locations within said video are determined based upon the content of the frame itself.

23 (previously presented): The method of claim 22 wherein said encoding is derived in accordance with the following:

$$Q/S_1, Q[,]/S_2, Q/S_3,..,Q/S_N$$

where Q is representative of said base quantization factor, and S_1 through S_N are representative of said sensitivity information for said plurality of locations.

24 (currently amended): ~~The method of claim 23 wherein one of said S_k, where k is a value from 1 to N, is derived based upon a statistical calculation of a plurality of said sensitivity information for one of said locations of said image.~~ A method of encoding a video comprising the steps of:

(a) receiving a frame of said video consisting of a plurality of pixels;

(b) calculating sensitivity information for a plurality of locations within said frame of said video calculated based upon the sensitivity of a human visual system of a viewer perceiving image detail at eccentric visual angles of a particular region of said frame of said video, where said particular region of said frame is determined based upon the content of the frame itself; and

(c) encoding said frame in a manner that provides a substantially uniform apparent quality to perceiving detail at eccentric visual angles of said plurality of locations of said frame to said viewer when said viewer is observing said particular region of said frame of said video, wherein said encoding of each of said plurality of locations of said frame of said video is based on a respective quantization value representative of a base

quantization factor divided by said sensitivity information for a respective one of said plurality of locations in a manner that said encoding employs at least two different quantization values, where said plurality of locations within said video are determined based upon the content of the frame itself, said encoding being derived in accordance with the following:

$Q/S_1, Q[.]/S_2, Q/S_3, \ldots Q/S_N$

where Q is representative of said base quantization factor, and S_1 through S_N are representative of said sensitivity information for said plurality of locations, wherein one of said S_k, where k is a value from 1 to N, is derived based upon a statistical calculation of a plurality of said sensitivity information for one of said locations of said image.

25 (previously presented): The method of claim 24 wherein S_k is an average of said plurality of said sensitivity information.

26 (previously presented): The method of claim 21 wherein said encoding of said frame of said video includes at least two different quantization values.

27 (previously presented): The method of claim 21 wherein said encoding said frame results in the total number of bits produced for said frame being substantially equal to a preselected number.

28 (previously presented): The method of claim 27 wherein said frame is encoded only once.

29 (previously presented): The method of claim 27 wherein said encoding of each of said plurality of locations is based on a respective quantization value representative of a base quantization factor divided by said sensitivity information for a respective one of said plurality of locations.

30 (previously presented): The method of claim 29 wherein said base quantization factor is derived in accordance with the following:

$$Q' = \sqrt{\frac{AK}{B - ANC}\sum_{i=1}^{N} \sigma_i^2 S_i^2}$$

where A is representative of the number of pixels in one of said plurality of locations, K and C are constants associated with said plurality of locations, N is representative of the number of said plurality of locations, B is representative of said total number of bits, the σ_i^2 values are a measure how much texture is associated with said plurality of locations, and the S_i^2 values are representative of the respective said sensitivity information squared.

31 (previously presented): A method for encoding multiple blocks in a frame of image data, comprising:

(a) identifying a target bit value equal to a total number of bits available for encoding the frame;

(b) calculating sensitivity information for each one of the blocks based upon the sensitivity of a human visual system perceiving image detail at eccentric visual angles of a particular region of the image, where said eccentricity of said particular region of said image is determined based upon the content of the frame itself;

(c) adapting quantization values for each of the multiple blocks to provide substantially uniform apparent quality to perceiving detail at eccentric visual angles of each of the blocks in the frame subject to a constraint that the total number of bits available for encoding the frame is equal to the target bit value; and

(d) encoding the blocks with the quantization values.

32 (previously presented): The method of claim 31 wherein the quantization values are derived according to the following,

$$Q' = \sqrt{\frac{AK}{B - ANC}\sum_{i-1}^{N} \sigma_i^{\,2} S_i^{\,2}}$$

where, Q_i is the quantization value for each block i, N is the number of blocks in the frame, B is the total number of bits available for encoding the frame, A is a number of pixels in each of the multiple blocks, K and C are constants associated with the image blocks, σ_i^2 is an empirical standard deviation of pixel values in the block, and S_i is a weighting incorporating the sensitivity information for the block.

33 (previously presented): The method of claim 31 including adjusting the quantization values according to a number of image blocks remaining to be encoded, a number of bits still available for encoding the remaining image blocks, and a value that depends on the sensitivity and texture of the remaining image blocks.

34 (previously presented): The method of claim 32 including using a K parameter and a C parameter on a block-by-block basis to adjust the quantization values for each of the multiple blocks, the K parameter modeling correlation statistics of the pixels in the image blocks and the C parameter modeling bits required to code overhead data.

35 (previously presented): The method of claim 34 including deriving the optimum quantization values in either a fixed mode where the K and C parameters are known in advance or an adaptive mode where the K and C parameters are derived according to the K and C parameters of previously encoded blocks.

36 (previously presented): The method of claim 35 wherein the adaptive mode includes the following steps:

(a) deriving values for the K and C parameters that exactly predict the number of bits B used for encoding previous blocks;

(b) deriving averages for the derived K and C parameters for the previously

encoded video blocks; and

(c) predicting the K and C parameters for a next video block by weighting the average K and C parameters according to the initial estimates for the K and C parameters.

37 (currently amended). A method for encoding video comprising the steps of:

(a) detecting the location of a facial region of a frame of said video;

(b) calculating a sensitivity value for each of a plurality of locations within said frame of said video based upon said location of said facial region <u>wherein said sensitivity values are calculated based upon a non-temporal said location of said facial region, a non-temporal size of said facial region, and a non-linear model of the human visual system's ability to perceive image detail at eccentric visual angles</u>; and

(c) encoding said frame in manner that provides a substantially uniform apparent quality to perceiving detail at eccentric visual angles of said plurality of locations to said viewer when said viewer is observing said facial region of said video.

38 (cancelled).

39 (previously presented): The method of claim 37 wherein said detecting said location of said facial region of said frame further comprises the steps of:

(a) receiving a first frame of said video comprising a plurality of pixels;

(b) receiving a subsequent frame of said video comprising a plurality of pixels;

(c) calculating a difference image representative of the difference between a plurality of said pixels of said first frame and a plurality of said pixels of said subsequent frame;

(d) determining a plurality of candidate regions within said difference image; and

(e) fitting said plurality of candidate regions to said difference image to select said facial region.

40 (previously presented): The method of claim 39 wherein said determining said plurality of candidate regions is based on a Hough transform of said difference image in a spacial domain to a parameter space.

41 (previously presented): The method of claim 39 further comprising the step of thresholding said difference image thereby removing values of said difference image less than a threshold value.

42 (previously presented): The method of claim 41 wherein said threshold value is a predetermined value and said removing values is setting said values less than said threshold value to a selected value.

43 (previously presented): ~~The method of claim 40 wherein said Hough transform is~~ ~~$A(x_c, y_c, r) = A(x_c, y_c, r) + 1 \; \forall x_c, y_c, r \in (x-x_c)^2 + (y-y_c)^2 = r^2$.~~ A method for encoding video comprising the steps of:

(a) detecting the location of a facial region of a frame of said video comprising the steps of:

(1) receiving a first frame of said video comprising a plurality of pixels;

(2) receiving a subsequent frame of said video comprising a plurality of pixels;

(3) calculating a difference image representative of the difference between a plurality of said pixels of said first frame and a plurality of said pixels of said subsequent frame;

(4) determining a plurality of candidate regions within said difference image wherein said determining said plurality of candidate regions is based on a Hough transform of said difference image in a spacial

domain to a parameter space wherein said Hough transform is $A(x_c, y_c, r) = A(x_c, y_c, r) + 1 \; \forall \; x_c, y_o, r \; \varepsilon \; (x-x_o)^2 + (y-y_c)^2 = r^2$; and

(5) fitting said plurality of candidate regions to said difference image to select said facial region;

(b) calculating a sensitivity value for each of a plurality of locations within said frame of said video based upon said location of said facial region; and

(c) encoding said frame in manner that provides a substantially uniform apparent quality to perceiving detail at eccentric visual angles of said plurality of locations to said viewer when said viewer is observing said facial region of said video.

44 (previously presented): The method of claim 39 where said fitting of each of said candidate regions is based on a combination of at least two of the following three factors including, a fit factor representative of a fit of said candidate regions to said difference image, a location factor representative of the location of said candidate regions within said video, and a size factor representative of the size of said candidate regions.

45 (previously presented): The method of claim 39 further comprising the step of scaling said first frame and said subsequent frame of said video to reduce the number of said pixels of said first and subsequent frame prior to said calculating said difference frame.

46 (previously presented): The method of claim 39 wherein said step of determining said plurality of said candidate regions and fitting said plurality of said candidate regions further comprises the steps of:

(a) determining a set of candidate circles based on a Hough transform of said difference image;

(b) scoring said set of said candidate circles based on a combination of at least three factors including, a fit factor representative of the fit of said candidate circles to said difference image, a location factor representative of the

location of said candidate circles within said video, and a size factor representative of the size of said candidate circles;

(c) selecting at least one of said candidate circles based on said scoring;

(d) generating at least one candidate region having an elliptical shape for each of said at least one of said candidate circles; and

(e) scoring each of said candidate regions based on a combination of at least three factors including, a fit factor representative of the fit of a respective said candidate region to said difference image, a location factor representative of the location of said respective said candidate region within said video, and a size factor representative of the size of said respective said candidate region.

47 (previously presented): The method of claim 46 wherein said generating at least one candidate region has a center of said elliptical shape located within a bounded region of potential locations having a greater vertical dimension than a horizontal dimension centered about the center of said respective said candidate circle.

48 (previously presented): The method of claim ~~38~~ 37 wherein said wherein the step of said calculating said sensitivity values is further based upon calculating an eccentricity versus image location in relation to a viewer of said video for said plurality of locations within said video.

49 (currently amended): The method of claim 48 wherein said eccentricity is derived according to the following,

$$\theta_E = \frac{180}{\pi} \tan^{-1}\left(\frac{\sqrt{\left(\frac{y - y_c}{y_R}\right)^2 + \left(\frac{x - x_c}{x_{[[i]]R}}\right)^2} - 1}{V} \right)$$

where θ_E is said eccentricity, y is a vertical pixel position within said video, x is a horizontal position within said video, x_c represents a horizontal component of a center position of an elliptical said facial region, y_c represents a vertical component of said center position of said elliptical said facial region, x_r represents a first elliptical radii of said elliptical said facial feature in a horizontal direction; y_r represents a second elliptical radii of said elliptical said facial feature in a vertical direction, and V represents a viewing distance.

50 (currently amended): The method of claim ~~38~~ 37 wherein said sensitivity values are based upon the distance from the outer edge of said facial region to said plurality of locations within said video.

51 (currently amended): The method of claim ~~38~~ 37 wherein said sensitivity versus eccentricity non-linear model is derived according to the following,

$$S = \frac{1}{1 + k_{ECC}\theta_E}$$

where S is representative of said sensitivity, k_{ECC} is a constant, and θ_E is representative of a non-linear contrast sensitivity function.

52 (previously presented): The method of claim 37 further comprising the step of encoding said frame in manner that provides a substantially uniform apparent quality of said plurality of locations to said viewer when said viewer is observing said facial region of said video.

53 (previously presented): The method of claim 37 wherein said encoding of each of said plurality of locations is based on a respective quantization value representative of a base quantization factor divided by said sensitivity information for a respective one of said plurality of locations.

54 (previously presented): The method of claim 53 wherein said encoding is derived in

accordance with the following:

$$Q/S_1, Q[,]/S_2, Q/S_3, .., Q/S_N$$

where Q is representative of said base quantization factor, and S_1 through S_N are representative of said sensitivity information for said plurality of locations.

55 (previously presented): The method of claim 54 wherein one of said S_k, where k is a value from 1 to N, is derived based upon a statistical calculation of a plurality of said sensitivity information for one of said locations of said image.

56 (previously presented): The method of claim 55 wherein S_k is an average of said plurality of said sensitivity information.

57 (previously presented): The method of claim 55 wherein S_k is a maximum of said plurality of said sensitivity information.

58 (previously presented): The method of claim 52 wherein said encoding said frame results in the total number of bits produced for said frame being substantially equal to a preselected number.

59 (previously presented): The method of claim 58 wherein said frame is encoded only once.

60 (previously presented): The method of claim 58 wherein said encoding of each of said plurality of locations is based on a respective quantization value representative of a base quantization factor divided by said sensitivity information for a respective one of said plurality of locations.

61 (previously presented): The method of claim 60 wherein said base quantization factor is derived in accordance with the following:

$$Q' = \sqrt{\frac{AK}{B-ANC}\sum_{i=1}^{N}\sigma_i^{\,2}S_i^{\,2}}$$

where A is representative of the number of pixels in one of said plurality of locations, K and C are constants associated with said plurality of locations, N is representative of the number of said plurality of locations, B is representative of said total number of bits, the σ_i^2 values are a measure how much texture is associated with said plurality of locations, and the S_i^2 values are representative of the respective said sensitivity information squared.